United Nations E/CN.3/2004/27

Economic and Social Council

Distr.: General
17 December 2003

Original: English

Statistical Commission
Thirty-fifth session
2-5 March 2004
Item 6 (f) of the provisional agenda*
Activities not classified by field: common open standards for the exchange and sharing of data and metadata

Report of the task force to establish standards on data and metadata exchange

Note by the Secretary-General

In accordance with a request of the Statistical Commission at its thirty-fourth session,** the Secretary-General has the honour to transmit the report of the task force to establish standards on data and metadata exchange. The Commission is invited to consider the points for discussion contained in paragraph 21 of the report.

* E/CN.3/2004/1.
** *Official Records of the Economic and Social Council, 2003, Supplement No. 4* (E/2003/24), chap. I.A.

03-66630 (E) 120104

Contents

		Paragraphs	*Page*
I.	Introduction	1–3	3
II.	The first phase of SDMX (2001-2003)	4–11	3
III.	The way forward (2004-2005)	12–18	5
IV.	Conclusions	19–21	7

Report on Common Open Standards for the Exchange and Sharing of Socio-economic Data and Metadata: the SDMX initiative[1]

I. Introduction

1. In 2001, the Bank for International Settlements (BIS), the European Central Bank (ECB), the Statistical Office of European Communities (Eurostat), the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD) and the United Nations joined together to develop more efficient processes for the exchange and sharing of data and metadata within the current scope of their collective activities. The World Bank joined the initial group of sponsor organizations in 2003.

2. The goal of the SDMX initiative is to explore common e-standards and ongoing standardization activities that could allow national and international organizations to gain efficiencies and avoid duplication of work. The above sponsor organizations intend to do this by taking advantage of existing and emerging exchange protocols, dissemination formats and e-standards.

3. The United Nations Statistical Commission received an initial report on the initiative during its thirty-third session, in March 2002.[2] A follow-up report was submitted to the Commission at its thirty-fourth session, in March 2003,[3] focusing on the launch of projects jointly undertaken by the sponsoring institutions. The present report reviews the latest progress in SDMX initiatives and outlines future plans. In particular, chapter II summarizes the main aims and achievements of the four projects launched in 2002. Chapter III presents the steps envisaged to be taken over the next two years and the expected implications for the sponsor organizations and the wider community of statistical data and metadata providers and users.

II. The first phase of SDMX (2001-2003)

4. The scope of the SDMX initiative is, in general terms, the development and maintenance of standards for statistical data and metadata exchange. These standards should be based on developments in information and communication technologies (ICT) and would be used in the context of a new cooperative effort on the part of international organizations to accomplish the following:

(a) Minimize the duplication of data and metadata collections;

(b) Maximize the coherence of data sets disseminated by various international organizations on the same subject matter;

(c) Improve the accessibility and interpretability of statistics disseminated by international organizations;

(d) Automate the data exchange processes and gain efficiency from the use of up-to-date ICT tools.

5. In particular, the SDMX initiative is focused on the exchange of data and metadata for economic and social statistics based both on the direct exchange of files between parties who have made prior agreements for the exchange, and on the

use of Internet technologies, including open exchange standards to make data and metadata accessible to all interested parties. The international organizations sponsoring the SDMX initiative are endeavouring to maintain existing standards and to create new ones that will suit not only their own needs but also those of their constituencies and users around the world.

6. From a technical point of view, the SDMX initiative has been looking at the development of a common information model that can foster the use of emerging solutions and technologies. In particular, this concerns the following:

(a) Electronic data interchange (EDI) conformant with the International Organization for Standardization (ISO) and the Centre for Facilitation of Procedures and Practices for Administration, Commerce and Transport (CEFACT) norms;

(b) The eXtensive Mark-up Language (XML) specified by the World Wide Web Consortium;

(c) Opportunities for web-based services that build on standards now being developed.

In addition, work has been undertaken to develop metadata standards.

7. Several activities have been developed in the first biennium of the initiative, under the chairmanship of Ms. C. Carson, Director of the Statistics Department of IMF. Four projects were established in June 2002:

- A practical case study on emerging e-standards for data and metadata exchange;
- The maintenance and advancement of standards for batch time series data exchange;
- The creation of a common vocabulary for statistical metadata;
- The development of a model for metadata repositories.

In addition, a web site has been developed and several presentations have been given to various audiences (such as, heads of national statistical offices, central bankers, ICT experts).

8. The aim of the practical case study was to investigate the ways in which emerging web standards can be used to simplify the collection, compilation and dissemination of statistical information. The case study focused on the Joint BIS-IMF-OECD-World Bank external debt statistics and national debtor data. Key deliverables included the following:

(a) Identifying e-standards that can foster the achievement of practical benefits;

(b) Demonstrating how web-based technologies may point to possible improvements and advances in statistical information exchange processes;

(c) Recommending ways to go forward with future SDMX activities.

The main documents from the case study and the practical demonstration are available on the SDMX web site (www.sdmx.org). The study revealed how new techniques and processes might work and illustrated potential benefits of emerging e-standards: opportunities for reduced reporting burden for suppliers, more user-

friendly access to national and international data published on the Web, easier comparative data analysis, and possible reuse of SDMX models and tools for statistical information systems at the national and international levels.

9. The second project focused on the release of the generic statistical message/time series (GESMES/TS) for data and metadata exchange in the batch EDI mode, a standard that has been used worldwide for more that five years in automating statistical data exchange. During 2003, in the light of the expected use of the message by a wide range of national and international organizations, its name was changed from GESMES/CB to GESMES/TS. Moreover, after extensive consultations between the project team and interested users with additional requirements, the updated release 3.0 of GESMES/TS was made available in July 2003. Finally, the preparation of an XML version of GESMES/TS 3.0 is under way.

10. The aim of the third project was the creation of a Metadata Common Vocabulary (MCV), containing a core set of metadata items and their related definitions to improve the standardization of metadata content for the purposes of data exchange and to promote the use of common nomenclatures that can foster international comparability of international data. The current version of the MCV (available on the SDMX web site) contains five fields — *term*, *definition*, *source*, *related terms* and *context* (the last one being used to provide additional explanation and to highlight some peculiarities) — and is intended to be of use to international organizations and national statistical agencies.

11. The fourth project was aimed at developing a standard approach to the creation and use of metadata repositories, allowing the user to access, analyse and reuse statistical metadata stored on multiple web sites. The project took as its starting point the Special Data Dissemination Standard (SDDS) framework from IMF. The main deliverables of the project are a metamodel for metadata repositories derived by adapting and expanding the SDDS framework to accommodate the needs of other SDMX partners and an XML schema and protocol to enable search and query across metadata repositories. This project has been closely coordinated with the MCV project.

III. The way forward (2004-2005)

12. Even though some technical development and maintenance work is still to be completed, the first phase of the SDMX work is coming to an end. Therefore, the sponsors discussed in the second half of 2003 how to continue the overall initiative, with the aim of delivering concrete solutions and starting the implementation of these solutions in several statistical subject matter domains. In addition to the technical work, a communication strategy has been designed to make potential users aware of SDMX proposals and solutions, and to involve national and international data providers in the change of current data and metadata exchange arrangements. Finally, the seven sponsors agreed to work over the biennium 2004-2005 under the chairmanship of Enrico Giovannini, Chief Statistician of OECD.

13. The first target of SDMX over the next biennium will be the release of version 1.0 standards for the SDMX Common Information Model, Data Formats and Core Metadata. In particular, the work on version 1.0 involves the key standards to support data and metadata sharing and web dissemination. These elements will also be incorporated in a start-up toolkit which will become freely available to all

interested users. Sponsors proposed to draft version 1.0 standards using a learn-by-doing approach. "External debt statistics" were identified as the most suitable subject matter for formulating such standards as an integral part of steps to develop a pilot implementation. In fact, this domain involves both international organizations (BIS-IMF-OECD-World Bank) and national agencies in producing and disseminating statistical data (respectively, from the creditor and the debtor perspectives) concerning the same phenomenon. The version 1.0 of SDMX standards will benefit from all results already achieved by the four projects carried out over the last biennium.

14. The Inter-Agency Task Force on Finance Statistics, at its September 2003 meeting, agreed to redesign the joint external debt statistics, integrating the currently available creditor sources of BIS, IMF, OECD and the World Bank with new statistics being released under the IMF Compilation Guidelines by debtor countries themselves. The intention is to extend the prototype developed by the SDMX case study to build a dynamic joint web hub linking all four organizations and national agencies producing and disseminating debt statistics.

15. Inter-Agency Task Force and SDMX members have agreed to use the resources for the joint external debt project to deliver version 1.0 standards and related tools, which will be generic to all statistical domains. A project proposal has been prepared to address the following deliverables:

(a) The first draft of version 1.0 SDMX core standards for data exchange and web dissemination and their release for public comment;

(b) The endorsement of version 1.0 core standards by the Sponsors Committee and steps to anchor them within the international standards community;

(c) Draft version 1.0 SDMX technical architecture standards and their release for public comment;

(d) The endorsement of version 1.0 of standards for technical architecture by sponsors and steps for anchoring them within the international standards community;

(e) The release of a pilot production infrastructure and of the associated External Debt Joint Hub implementation.

16. In addition, SDMX sponsors will start preparing the groundwork for implementing the SDMX standards in other statistical domains. OECD is working on the adoption of SDMX standards to share national accounts data with member countries and other international organizations. OECD and IMF are exploring the use of web services between internal database systems at these institutions. For international trade data, OECD and the Statistics Division are building a shared database based on SDMX standards. Finally, at the beginning of 2004, SDMX sponsors will identify one or two additional subject matter domains in which to implement data and metadata exchange. The idea is that the implementation of SDMX in these and other statistical domains will promote the future development of new versions of SDMX standards.

17. Apart from developing a sequenced implementation of SDMX standards in various statistical domains, the SDMX sponsors also envisage to foster the adoption of GESMES/TS (for example, via training programmes) and the regular updating of

the MCV. Moreover, they envisage the release of a production version of the metamodel for metadata repositories based on SDMX standards.

18. As it is not possible to develop world standards on statistical data and metadata exchange without having close relationships with other bodies in charge of setting similar standards and with ultimate potential users of the latter, special attention will be paid over the next two years to the communication of SDMX results to potential users and partners. Relationships with national statistical data providers (such as national statistical offices and central banks) and international organizations active in statistics will be reinforced, using existing bodies belonging to SDMX sponsor organizations, and the SDMX web site will be redesigned. In addition, relationships with international bodies (for example, ISO) which set world standards linked to ICT matters, international information brokers and groups involved in information policy management, such as groups focusing on e-government standards (for example, the Ozon Action Strategic Information System) or accounting standards (for example, the International Federation of Accountants), will be established or reinforced.

IV. Conclusions

19. The SDMX initiative has the potential to make an important contribution in helping international and national agencies involved in the production of statistical data and metadata to become more efficient in their internal operations, to avoid duplication in data collection and to provide users with a higher-quality service. The first set of complete results of the initiative is expected in 2005, even if important deliverables (such as GESMES/TS 3.0, the Metadata Common Vocabulary) are already available to users and/or will be ready for external distribution over the next 12 months.

20. New ICT developments can be applied to several aspects of statistical activities, but to fully benefit from them, international and national organizations must also address governance issues. In some cases, SDMX standards may provide a possibility to statistical "communities" to reconsider the current pattern of data and metadata exchange and change the existing division of labour. On the other hand, new models for data exchange and dissemination based on the concept of "hub" (a common registry keeping track of data and metadata available on a network of web sites of partners) can reinforce the role of national statistical offices in coordinating other data providers. Finally, international organizations have the opportunity to avoid duplication in data collection, share data and metadata among themselves and with other users, and establish new ways to co-manage international statistical activities and to cooperate in that regard.

21. **The United Nations Statistical Commission may wish to comment on the expected technical developments and to discuss how these can impact data exchange practices and produce changes in the overall governance of national and international statistical systems**.

Notes

[1] The report has been jointly prepared by the Bank for International Settlements, the European Central Bank, the Statistical Office of European Communities, the International Monetary Fund, the Organisation for Economic Cooperation and Development, the United Nations Statistics Division and the World Bank.

[2] See *Official Records of the Economic and Social Council, 2002, Supplement No. 4* (E/2002/24), chap. VI, sect. C. See also http://unstats.un.org/unsd/statcom/sc2002.htm.

[3] See http://unstats.un.org/unsd/statcom/sc2003.htm.